WEST

Help | Logout | Interrupt

Main Menu | Search Form | Posting Counts | Show S Numbers | Edit S Numbers | Preferences | Cases

Search Results -

Terms	Documents
l2 and l3 and l4	3

Database:
US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L7

Refine Search

Recall Text | Clear

Search History

DATE: Wednesday, August 20, 2003 Printable Copy Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=TDBD; PLUR=YES; OP=OR		
L7	l2 and l3 and l4	3	L7
L6	l2 and l3 and l4 and L5	1	L6
L5	euclidean$	80	L5
L4	distance$	4581	L4
L3	(node$ or leave$ or leaf$) near5 (split$ or partition$ or divid$ or division or segment$ or fragment$ or pars$ or break$ or broken or separat$ or chop$)	261	L3
L2	decision adj tree$	63	L2
L1	tree$ or hierarch$	2264	L1

END OF SEARCH HISTORY

Search Results - Record(s) 1 through 3 of 3 returned.

☐ 1. Document ID: NN9312157

L7: Entry 1 of 3 File: TDBD Dec 1, 1993

TDB-ACC-NO: NN9312157

DISCLOSURE TITLE: Construction of Context Dependent Label Prototypes for Use in a Speech Recognition System

PUBLICATION-DATA:
IBM Technical Disclosure Bulletin, December 1993, US

VOLUME NUMBER: 36
ISSUE NUMBER: 12
PAGE NUMBER: 157 - 158

PUBLICATION-DATE: December 1, 1993 (19931201)

CROSS REFERENCE: 0018-8689-36-12-157

DISCLOSURE TEXT:

In discrete-parameter speech recognition systems, a vector quantiser outputs an acoustic label at regular intervals. In one prominent approach to speech recognition 1ù, each label is characterized by a "prototype" consisting of a mixture of diagonal Gaussian distributions, and the label output identifies the prototype which maximizes the likelihood of a corresponding acoustic parameter vector. There is one prototype mixture per label, and it does not depend on the phonetic context of the frame being labelled. - The invention below generalizes the concept of Gaussian mixture prototypes so as to make them context-dependent. Instead of having one mixture per label, there are several mixtures per label, one of which will be selected to assess the likelihood of an acoustic vector. The appropriate mixture is determined from the phonetic context of the corresponding frame. - The idea of context-dependent acoustic modelling is not new: it forms the basis of the acoustic Markov word models previously described in 2ù. But in 2ù it is the word models that are context dependent, not the label prototypes as advocated here. - Others have also recently suggested the use of context-dependent prototypes 3ù. The invention below differs from 3ù in the following principal ways. First, we advocate different acoustic parameters. Second, we obtain context-dependency rules via decision trees which maximize Gaussian likelihoods, whereas in 3ù they seek to minimize Euclidean distances. Third, decision trees are constructed for each vector element separately in 3ù, rather than a single decision tree for the entire vector as here. Fourth, the prototypes in 3ù consist of single Gaussians. The present invention uses a mixture of diagonal Gaussians. Fifth, in 3ù the context-dependency rules cover only one phoneme on each side of the frame being modelled, whereas the method below covers several on each side: typically five. - Assume that some training data has been recorded, signal processed, and Viterbi aligned against phoneme-based Markov word models as described in 1,4ù. - Assume further, that the existence of some phonetically meaningful questions which may be sued to construct phonological trees. These questions, which may be applied to any phone P in the neighborhood of the frame being processed, usually take the form "is P a member of the set S?". Here S denotes a set containing one or more phonetic phones having something in common. The necessary sets may be obtained from almost any phonetic text book. For present purposes, "word boundary" is also considered to be a phonetic phone. The following steps are performed. 1. Using the existing Viterbi alignments, tag each parameter vector in the training data with (1) the identity of the arc against which the vector was aligned, (2) the phonetic phone which contained that arc, (3) the N phonetic phones which preceded that phone, and (4) the N phonetic phones which

followed it. A typical value for N is 5. 2. Perform Steps 3-6 for each arc A in the arc inventory. 3. Extract from the tagged data of Step 1, all the data which was aligned with arc A. 4. Construct a phonological tree from the data extracted in Step 3, so as to maximize the joint likelihood of the acoustic vectors when modelled by a separate diagonal Gaussian at each leaf. The details of tree construction are given in 5ù. At each node of this tree there is a binary question relating to one of the (2N + 1) phonetic phones extracted in Step 1. These questions are selected from the set of phonetically meaningful questions discussed above. 5. Perform Step 6 at each leaf of the tree for arc A. 6. Cluster the parameter vectors associated with leaf A into K diagonal Gaussians as described in 1ù Reasonable values for K lie in the range 2-10. K = 1 is not a good choice. - At the completion of Step 6, a phonological tree will have been constructed for each arc in the arc inventory. At each leaf L of the phonological tree for any given arc A, there is a mixture of diagonal Gaussian distributions. This mixture characterizes the acoustic parameter vectors associated with arc A when the phonetic context of arc A satisfies the conditions which lead to leaf L. Thus we have created a set of context-dependent prototypes for each arc in the arc inventory. Defining the label alphabet to be same as the arc inventory yields a set of context-dependent label prototypes in the form of a mixture of diagonal Gaussian distributions. - The acoustic parameter vectors referenced above are created from a window of M consecutive frames as described in 1,4ù where M is typically about 9. Other windows such as those of 6,7ù may be used instead. References 1ù U.S. Patent 5,182,773. 2ù U.S. Patent 5,033,087. 3ù M. Phillips, J. Glass and V. Zue, "Modelling Context Dependency in Acoustic-Phonetic and Lexical Representations. Proc. of Fourth DARPA Workshop on Speech and Natural Language (1991). 4ù "Construction of Markov Word Models for Computer Recognition of Continuous Speech," IBM Disclosure Technical Disclosure Bulletin 36, 11 (November 1993). 5ù "Construction of a Tree for Context-Dependent prototypes in a Speech Recognition System," IBM Disclosure Bulletin 36, 12 (December 1993). 6ù "Construction of a Projection Matrix Spanning a Large Time Window for use in a Speech Recognition System," IBM Technical Disclosure Bulletin 36, 9A (September 1993).

7ù "Method for Constructing a Sparse Projection Matrix Spanning a Large Time Window for use in a Speech Recognition System," IBM Technical Disclosure Bulletin 36, 9A (September 1993).

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw Desc

---

☐ 2. Document ID: NN9108217

L7: Entry 2 of 3 File: TDBD Aug 1, 1991

TDB-ACC-NO: NN9108217

DISCLOSURE TITLE: New Varieties of Fenemic Markov Models for Continuous Speech Recognition.

PUBLICATION-DATA:
IBM Technical Disclosure Bulletin, August 1991, US

VOLUME NUMBER: 34
ISSUE NUMBER: 3
PAGE NUMBER: 217 - 218

PUBLICATION-DATE: August 1, 1991 (19910801)

CROSS REFERENCE: 0018-8689-34-3-217

DISCLOSURE TEXT:

- In continuous speech the co-articulation effect changes the pronunciation of words considerably. An effective way of capturing these phenomena automatically based on constructing decision trees is described in been*|. In this method the label sequences from several utterances of each phone in different contexts are extracted. A decision tree is built for each phone by interrogating the context in which the phone occurs. The goodness of each split is measured by fitting a model to the strings at each node and measuring how much improvement in the fit is obtained by the splits. The label strings that end up at one particular leaf in the tree are used to make a fenemic baseform for the phone in the context given by the answers to the questions leading to that leaf from the root of tree. These baseforms are made by using a variety of 200 or so fenemic models and determining the fenemic machine sequence that maximizes the joint likelihood of the strings at the leaf. - It is likely that 200 or so fenemic machines are not enough to accurately model all the sounds occurring in continuous speech. The invention described here provides a method for increasing the number of fenemic phone varieties so as to make the models more accurate. - The method used for determining the necessary variety of fenemic models is the following: 1. Construct the decision trees and the fenemic baseforms using the method described in been*| and the original fenemic models. Let the original fenemic models be numbered f1f2 ...,fF and the underlying phones be denoted by P1P2 ...,Pn . 2. Align the training data from several speakers (say, 10) against the baseforms constructed in step 1 using the Viterbi algorithm. This gives the fenemic model and the leaf of the tree against which each label in the training data is aligned. 3. For each pair of underlying phone Pi and original fenemic machine fj compute the count for each label. Determine the output probability distribution for the fenemic model fj occurring in the baseforms for the underlying phone Pi from these counts. Let these distributions be denoted by dij . 4. For each j cluster the distributions dij using a bottom-up clustering algorithm and an appropriate distance measure. A possible distance is given by Only the distributions for the same original fenemic machine occurring in the baseforms of different underlying phones Pi are clustered in this manner. The clustering is repeated separately for each fenemic machine fj . The clustering is controlled by a maximum distance parameter. Two clusters are merged only if the distance between them falls within this limit. This parameter can be chosen so that varieties of distributions that differ significantly are identified. 5. Define a new fenemic model for each cluster left after the above step. This creates a mapping from the pair Pifj to the new fenemic model ck . 6. Map the baseforms produced in step 1 to the new fenemic models. Re-write each original fenemic model fj occurring in the baseforms for a phone Pi as a new machine ck using the mapping produced in the step above. - These are now the new baseforms to be used in the continuous speech recognition algorithms. The parameters of these models are trained in the usual manner. The initial values for the parameters can be obtained by looking at the original fenemic machine from which the new models were derived. - Reference (*) L. R. Bahl, P. V. deSouza, P. S. Gopalakrishnam, D. Nahamoo and M. A. Picheny "Decision Trees for Phonological Rules in Continuous Speech," Proceedings of the ICASSP-91 (1991).

3. Document ID: NB900666

L7: Entry 3 of 3 File: TDBD Jun 1, 1990

TDB-ACC-NO: NB900666

DISCLOSURE TITLE: Algorithm for Concept Learning Utilizing Novel Attribute Space-Dividing Method.

PUBLICATION-DATA:
IBM Technical Disclosure Bulletin, June 1990, US

VOLUME NUMBER: 33
ISSUE NUMBER: 1B
PAGE NUMBER: 66 - 68

PUBLICATION-DATE: June 1, 1990 (19900601)

CROSS REFERENCE: 0018-8689-33-1B-66

DISCLOSURE TEXT:

- Disclosed is an algorithm of "CONCEPT LEARNING" which utilizes a novel attribute space-dividing method. The following three techniques are invented for automatic generation of "A Binary Decision Tree" which is often utilized as a schema for concept learning: (1) automatic generation of new "USEFUL ATTRIBUTES",(2) automatic generation of "GENERALIZATION TREE", and (3) a new criterion for deciding the best decision formula. - Concept learning is a well known technique in which, by analyzing large number of data as to the relationships between a set of attributes and the class to which the data belongs, the generalized expression of relationships between each class and attributes is derived. This process is carried out mainly by dividing the attribute space (multi-dimensional). The derived expression is utilized for predicting which class a certain new data belongs to. A "BINARY DECISION TREE", such as shown in Fig. 1, is often utilized as a scheme for attribute space division. A decision formula is generated at each node, and by this formula, the data for learning are divided into two groups which, respectively, formulate another new node. This division process is iterated and the tree is extended until each tip node nearly consists of data belonging to only one class. - The conventional technique in this field can be used only under quite ideal conditions and is of little use in the practical application areas where empirical data are the mere source of learning. The invention overcomes these deficiencies by employing a new attribute space-dividing method and provides powerful concept learning capability even in the practical application area. The detailed descriptions of the newly invented techniques are as follows. - Method for automatic generation of new USEFUL ATTRIBUTES: By programming some basic mathematical operators (such as +, -, *, /, etc.) beforehand, various new attributes are defined automatically during the learning phase in the form of multivariable function of two basic attributes binded by one of operators. A new attribute is defined if a combination of two attributes yields sufficient improvement in the criterion for deciding the best decision formula. Recursive definition is also possible, which yields quite high flexibility to describe the cutting plane within the attribute space. - Method for automatic generation of GENERALIZATION TREE: When using discrete type attributes, it is difficult to formulate a binary decision tree because the mutual distance of attribute values are not known beforehand. First, the pair of attribute values which most frequently appears in the data of the same class is considered to belong to one group, and then these values are replaced by the automatically defined group name. This process is iterated until no such pairing is found. The generated hierarchy of group names constitutes the generalization tree. Fig. 2 is an example of the generated generalization tree. Two values appearing in the nearby positions of this tree are considered to be "of little distance" and vice versa. - New criterion for the decision formula at each node: This is a new criterion to determine which decision formula to take for dividing the data belonging to each node. First, the "occupancy ratio" of each node is defined as the ratio of the number of data of the majority class of this node to the number of total data of this node. The new criterion for deciding the best decision formula is to maximize the number of data of the positive node (the group whose data satisfy the decision formula) while keeping the occupancy ratio of the positive node above a certain value (80 or 90%). This new criterion is quite suited for the division of attribute space with inaccurate or noisy learning data. If the target occupancy ratio is set (at 80 or 90%, for example) before learning, the attribute space can be finally divided roughly keeping this level with this method.

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | KWIC
Draw Desc | Clip Img

Generate Collection | Print

Terms	Documents
l2 and l3 and l4	3

Display Format: FULL | Change Format

Previous Page   Next Page

STIC EIC 2100 101739
Search Request Form

Today's Date: 8.20.3

What date would you like to use to limit the search?
Priority Date: 7.30.01 Other:

Name J. FLEURANTIN
AU 2172 Examiner # 75536
Room # 4A50 Phone 308-6718
Serial # 9/918,952

Format for Search Results (Circle One):
PAPER DISK EMAIL

Where have you searched so far?
USP DWPI EPO JPO ACM IBM TDB
IEEE INSPEC SPI Other

Is this a "Fast & Focused" Search Request? (Circle One) YES NO
A "Fast & Focused" Search is completed in 2-3 hours (maximum). The search must be on a very specific topic and meet certain criteria. The criteria are posted in EIC2100 and on the EIC2100 NPL Web Page at http://ptoweb/patents/stic/stic-tc2100.htm.

What is the topic, novelty, motivation, utility, or other specific details defining the desired focus of this search? Please include the concepts, synonyms, keywords, acronyms, definitions, strategies, and anything else that helps to describe the topic. Please attach a copy of the abstract, background, brief summary, pertinent claims and any citations of relevant art you have found.

- decision tree
* leaf node
- distance function
* distance based on partitions
* gini index
  2 "gini index"

STIC Searcher Geoffrey St. Leger Phone 308-7800
Date picked up 4B30 Date Completed 8/20/3

STIC Search Report
EIC 2100

USPTO

STIC Database Tracking Number: 101739

TO: Jean Fleurantin
Location: 4A50
Art Unit : 2172
Wednesday, August 20, 2003

Case Serial Number: 09/918952

From: Geoffrey St. Leger
Location: EIC 2100
PK2-4B30
Phone: 308-7800

geoffrey.stleger@uspto.gov

Search Notes

Dear Examiner Fleurantin,

Attached please find the results of your Fast & Focused search request for application 09/918952. I searched Dialog's foreign patent files, technical databases, product announcement files and general files.

Please let me know if you have any questions.

Regards,

Geoffrey St. Leger

Geoffrey St. Leger
4B30/308-7800

STIC Search Results Feedback Form

USPTO

EIC 2100

Questions about the scope or the results of the search? Contact ***the EIC searcher or contact***:

Anne Hendrickson, EIC 2100 Team Leader
308-7831, CPK2-4B40

Voluntary Results Feedback Form

- *I am an examiner in Workgroup:* ☐ *Example: 3730*
- *Relevant prior art **found**, search results used as follows:*
  - ☐ 102 rejection
  - ☐ 103 rejection
  - ☐ Cited as being of interest.
  - ☐ Helped examiner better understand the invention.
  - ☐ Helped examiner better understand the state of the art in their technology.

  Types of relevant prior art found:
  - ☐ Foreign Patent(s)
  - ☐ Non-Patent Literature
    (journal articles, conference proceedings, new product announcements etc.)
- *Relevant prior art **not found:***
  - ☐ Results verified the lack of relevant prior art (helped determine patentability).
  - ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/EIC2100 CPK2-4B40

sira Search and Information Resources Administration

Search Results

Search Results for: **[euclidean distance<AND>((euclidean*<AND>((distance<AND>(((node* or leave* or leaf*) <near/5> (split* or partition* or divid* or division or segment* or fragment* or pars* or break* or broken or separat* or chop*)<AND>((decision tree*))))))))]**
Found **7** of **120,398 searched.**

Results 1 - 7 of 7 **short listing**

1 Scalable algorithms for mining large databases — 84%
Rajeev Rastogi , Kyuseok Shim
Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining August 1999

2 Identifying prospective customers — 82%
Paul B. Chou , Edna Grossman , Dimitrios Gunopulos , Pasumarti Kamesam
Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining August 2000

3 High performance data mining (tutorial PM-3) — 80%
Vipin Kumar , Mohammed Zaki
Tutorial notes of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining August 2000

4 Machine learning in DNA microarray analysis for cancer classification — 77%
Sung-Bae Cho , Hong-Hee Won
Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003 - Volume 19 January 2003

The development of microarray technology has supplied a large volume of data to many fields. In particular, it has been applied to prediction and diagnosis of cancer, so that it expectedly helps us to exactly predict and diagnose cancer. To precisely classify cancer we have to select genes related to cancer because extracted genes from microarray have many noises. In this paper, we attempt to explore many features and classifiers using three benchmark datasets to systematically evaluate the perf ...

5 Techniques for Reducing Pen Plotting Time 77%
D. P. Anderson
ACM Transactions on Graphics (TOG) July 1983
Volume 2 Issue 3

6 An Algorithm for Finding Best Matches in Logarithmic Expected Time 77%
Jerome H. Freidman , Jon Louis Bentley , Raphael Ari Finkel
ACM Transactions on Mathematical Software (TOMS) September 1977
Volume 3 Issue 3

7 Efficient algorithms for mining outliers from large data sets 77%
Sridhar Ramaswamy , Rajeev Rastogi , Kyuseok Shim
ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000
Volume 29 Issue 2

In this paper, we propose a novel formulation for distance-based *outliers* that is based on the distance of a point from its k^{th} nearest neighbor. We rank each point on the basis of its distance to its k^{th} nearest neighbor and declare the top n points in this ranking to be outliers. In addition to developing relatively straightforward solutions to finding such outliers based on the classical nested-loop join and index join algorithms, we develo ...

Results 1 - 7 of 7 **short listing**

Advanced Web Search

Search with...

this boolean expression

decision tree* AND ((node* or leave* or leaf*) NEAR (split* or partition* or divid* or division or segment* or fragment* or pars*

Basic Search

Use terms such as AND, OR, AND NOT, NEAR

sorted by

Pages with these words will be ranked highest.

SEARCH: ◉ Worldwide ○ U.S. **RESULTS IN:** ○ All languages ◉ English, Spanish

We found 0 results. Suggestions:

- Check your spelling.
- Try different or fewer keywords.
- Remove quotation marks or plus signs.

Extend Your Search About

Search the Yellow Pages for **euclidean decision "or" divi...** with SMARTpages.com

Advanced Web Search

○ **Build a query with...**

all of these words

this exact phrase

any of these words

and none of these words

B

◉ **Search with...**

this boolean expression

decision tree* AND ((node* or leave* or leaf*) NEAR (split* or partition* or divid* or division or segment* or fragment* or pars*

Use terms such as AND, AND NOT, NEAR

sorted by

Pages with these words will be ranked highest.

SEARCH: ◉ Worldwide ○ U.S. **RESULTS IN:** ○ All languages ◉ English, Spanish